# Unit Root Testing in Heteroskedastic Panels using the Cauchy Estimator*

Matei Demetrescu[†] Christoph Hanck[‡]

May 19, 2011

**Abstract**

The Cauchy estimator of an autoregressive root uses the sign of the first lag as instrumental variable. The resulting IV $t$-type statistic follows a standard normal limiting distribution under a unit root case even under unconditional heteroskedasticity, if the series to be tested has no deterministic trends. The standard normality of the Cauchy test is exploited to obtain a standard normal panel unit root test under cross-sectional dependence and time-varying volatility with an orthogonalization procedure. The paper's analysis of the joint $N, T$ asymptotics of the test suggests that (1) $N$ should be smaller than $T$ and (2) its local power is competitive with other popular tests. To render the test applicable when $N$ is comparable with, or larger than $T$, shrinkage estimators of the involved covariance matrix are used. The finite-sample performance of the discussed procedures is found to be satisfactory.

**Keywords:** Integrated process, Time-varying variance, Nonstationary volatility, Asymptotic normality, Cross-dependent panel, Joint asymptotics
**JEL classification:** C12 (Hypothesis Testing), C23 (Models with Panel Data)

## 1 Motivation

Instrumental variable [IV] estimation is typically used to deal with regressor endogeneity, but has turned out to be a valuable tool in unit root econometrics as well. So and Shin (1999) showed that the IV estimation procedure using the sign of the first lag as instruments for the lag itself has nice properties: the $t$ statistic based on this so-called Cauchy estimator has a standard normal limiting distribution under i.i.d. innovations and stationary, unit, or explosive roots in the examined series. In spite of standard asymptotics, the Cauchy test has nontrivial power in $T^{-1}$-neighborhoods of the unit root (Demetrescu and Hanck, 2011); and, unlike OLS based tests, the Cauchy test can easily be used in a nonlinear or seasonal time series framework (Shin and Lee, 2001, 2003).

But (near-)integration is not the only form of nonstationarity data can exhibit: the data often have time-varying variances even after taking logs. A prominent example is the so-called Great

*The authors would like to thank an anonymous referee and an associate editor, as well as Peter Boswijk, Jörg Breitung, Hashem Pesaran, and Werner Ploberger for very helpful comments and suggestions.

[†]Institute of Econometrics and Operations Research, Department of Economics, University of Bonn, Adenauerallee 24-42, D-53113 Bonn, Germany, tel. +49 228 733925, email: matei.demetrescu@uni-bonn.de.

[‡]Corresponding author. Rijksuniversiteit Groningen, Department of Economics, Econometrics and Finance, Nettelbosje 2, 9747AE Groningen, Netherlands, tel. +31 50 3633836, email: c.h.hanck@rug.nl.

Moderation, i.e. the decline in the volatility of many economic variables towards the end of the 1900's (Stock and Watson, 2002). Cavaliere (2004) shows that the null distribution of the ADF test then depends on nuisance parameters. In contrast, Demetrescu and Hanck (2011) show the Cauchy unit root test to be robust to such heteroskedasticity. And robustness to unconditional heteroskedasticity is relevant for panel unit root tests just like it is for univariate tests: we demonstrate in this paper that several popular second-generation panel unit root tests cease to work reliably under unconditional heteroskedasticity in the time dimension.

The paper therefore studies the asymptotic behavior of panel unit root tests based on the Cauchy estimator in panels with unconditionally heteroskedastic innovations as follows.

After briefly discussing the univariate case in Section 2, we establish in Section 3 standard normality of the orthogonalization procedure proposed by Shin and Kang (2006) under joint $N, T$-asymptotics. The cross-unit correlation is modeled by a factor structure of the errors, allowing for strong cross-correlation and time-varying variance. The admissible rates for $N$, however, are required to be slower than $T^{1/5}$, also because Shin and Kang's procedure requires orthogonalization with an estimated $N \times N$ covariance matrix. We also demonstrate the test to have power against local alternatives of the form $N^{-0.5}T^{-1}$.

Finite sample simulations in Section 4 confirm the asymptotic predictions. The size is well-controlled for cross-correlated panels exhibiting e.g. variance breaks at heterogenous times as long as $T$ is larger than $N$. We overcome this slight drawback by using shrinkage estimators of the covariance matrix such that the test works reliably for larger $N$. Alternatively, combining single-unit Cauchy statistics along the lines of Hartung (1999) leads to similarly reliable panel tests under heteroskedasticity.

## 2 The univariate Cauchy unit root test

We begin by giving the necessary univariate background. The data generating process [DGP] has the additive representation $y_t = m + x_t$, $t = 1, \ldots, T$, where $x_t = \rho x_{t-1} + u_t$, $x_0$ fixed, with possibly a unit root and $u_t$ a stable AR($p$) process. We refer to Demetrescu and Hanck (2011) for a more detailed discussion of other deterministic specifications and the assumptions as well as univariate simulation evidence. With $\phi = \rho - 1$, the unit root null is $\phi = 0$ in the representation

$$\Delta x_t = \phi x_{t-1} + \sum_{j=1}^{p} a_j \Delta x_{t-j} + \varepsilon_t. \tag{1}$$

Recursive demeaning is required, as the cross-product of instrument and $\varepsilon_t$ needs to be a martingale difference [md] (see So and Shin, 1999); so instrument $\widetilde{y}^{\mu}_{t-1} = y_{t-1} - \frac{1}{t-1}\sum_{j=1}^{t-1} y_j$ by $h\left(\widetilde{y}^{\mu}_{t-1}\right)$ in the regression

$$\Delta y_t = \widehat{\phi}\widetilde{y}^{\mu}_{t-1} + \sum_{j=1}^{p} \widehat{a}_j \Delta y_{t-j} + \widehat{\varepsilon}_t, \tag{2}$$

with $\Delta y_{t-j}$ instrumenting themselves and $h(\cdot)$ a Huber-type instrument (asymptotically equivalent to the sign) as in Shin and Kang (2006). The test statistic is $t^{\mu}_{IV} = \widehat{\phi}/s.e.\big(\widehat{\phi}\big)$. Following Cavaliere and Taylor (2007), the $\varepsilon_t$ are unconditionally heteroskedastic. But we relax their i.i.d. assumption:

**Assumption 1.** *Let $\varepsilon_t = \sigma_t \epsilon_t$, where $\sigma_t = \omega\,(t/T) > 0$ is bounded on $[-\infty; 1]$ and $\omega$ is piecewise Lipschitz, and $\epsilon_t$ is md with uniformly bounded conditional (on $\{\epsilon_{t-1}, \epsilon_{t-2}, \ldots\}$) densities such that $\mathrm{E}(\epsilon_t^2) = 1$, $\exists r > 4$ with $\sup_t \mathrm{E}\,|\epsilon_t|^r < \infty$, and $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{E}\big(\big|\tilde{\epsilon}_s^2 - 1\big|\big|\tilde{\epsilon}_t^2 - 1\big|\big) < \infty$.*

Under Assumption 1, $\frac{1}{\sqrt{T}}x_{[sT]} \Rightarrow \frac{(\int_0^1 \omega^2(r)\,\mathrm{d}r)^{1/2}}{1-\sum_{j=1}^{p} a_j} J_c\,(\eta(s))$, a time-transformed OU process, where $\eta\,(s) = \big(\int_0^1 \omega^2\,(r)\,\mathrm{d}r\big)^{-1}\int_0^s \omega^2\,(r)\,\mathrm{d}r$. The distribution of $t^{\mu}_{IV}$ is then given by

**Proposition 1.** *Under* (1), *local alternatives $\phi = -c/T$ with $c \geq 0$ and Assumption 1,*

$$t^{\mu}_{IV} \xrightarrow{d} \int_0^1 \operatorname{sgn}\left(\widetilde{J}^{\mu}_{c,\eta}(s)\right) \mathrm{d}W\,(\eta(s)) - \frac{c}{1-\sum_{j=1}^{p} a_j}\int_0^1 \operatorname{sgn}\left(\widetilde{J}^{\mu}_{c,\eta}(s)\right) J_c\,(\eta(s))\,\mathrm{d}s$$

*as $T \to \infty$, where $\widetilde{J}^{\mu}_{c,\eta}(s) = J_c\,(\eta(s)) - \frac{1}{s}\int_0^s J_c\,(\eta(r))\,\mathrm{d}r$ and $\widetilde{J}^{\mu}_{c,\eta}(0) = 0$ a.s. Under the null $c = 0$,*

$$t^{\mu}_{IV} \xrightarrow{d} \mathcal{N}(0,1).$$

**Proof:** See Demetrescu and Hanck (2011).

Intuitively, heteroskedasticity-robustness is obtained because sgn discounts the large variability of the lagged level to 1 or $-1$ irrespective of how the volatility process changes in $t$.

# 3 IV panel unit root tests

Let $y_{i,t}$ be the observed panel, generated as $y_{i,t} = m_i + x_{i,t}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$. The stochastic component $x_{i,t}$ is a unit-wise autoregression of order $p_i + 1$ with a possible unit root:

$$\Delta x_{i,t} = \phi_i x_{i,t-1} + \sum_{j=1}^{p_i} a_{ij}\Delta x_{i,t-j} + \varepsilon_{i,t} \tag{3}$$

with uniformly (in $i$) bounded starting values. Under the unit root null, $\rho_i = 1$ or $\phi_i = 0$. We use for simplicity $h_i(\cdot) \equiv h(\cdot)\,\forall\, i$ but allow the unit-specific DGPs to exhibit heterogenous $p_i$ with finite maximal order (set "missing" $a_{ij}$ in units with lower actual order to zero).

**Assumption 2.** *Let $\sup_i p_i \leqslant p$, $i = 1, \ldots, N$ for some $p$ not depending on $T$ or $N$.*

We demonstrate below that many popular second-generation panel unit root tests (e.g., Pesaran, 2007; Breitung and Das, 2005; Demetrescu et al., 2006; Moon and Perron, 2004) do not control size under unconditional heteroskedasticity in the time dimension. On the contrary, the Cauchy test's univariate robustness to unconditional heteroskedasticity prevents such failure in the panel case as well. Now, the test suggested in Demetrescu et al. (2006) combines individual ADF tests, and fails because of the ADF's lack of robustness to unconditional heteroskedasticity; when replacing ADF tests with Cauchy tests, the test works more reliably, see Section 4.

Under cross-sectional independence, panel tests can easily be built from the single-unit tests $t^{\mu}_{IV,i}$ due to their standard asymptotics; $1/\sqrt{N}\sum_{i=1}^{N} t^{\mu}_{IV,i}$ for instance yields a standard normal panel statistic. This holds when allowing for $N \to \infty$; but $N \to \infty$ is not necessary for normality.

Under cross-correlation, the Cauchy panel unit root test requires orthogonalization, since the individual test statistics are correlated (Shin and Kang, 2006). Let $\overline{\varepsilon}_{i,t} = \Delta y_{i,t} - \sum_{j=1}^{p} \overline{a}_{ij}\Delta y_{i,t-j}$ and $\overline{\boldsymbol{\varepsilon}}_t = (\overline{\varepsilon}_{1,t}, \ldots, \overline{\varepsilon}_{N,t})'$ be the prewhitened differences; as estimates $\overline{a}_{ij}$, Shin and Kang (2006) suggest using OLS estimates under the null $\rho_i = 1$. Then, compute the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\varepsilon} = \frac{1}{T-p}\sum_{t=p+2}^{T} \overline{\boldsymbol{\varepsilon}}_t \overline{\boldsymbol{\varepsilon}}_t'$ and let $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^{-1} = \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}}'$ be a suitable $LU$ decomposition. Denote the orthogonalized, prewhitened differences by $\boldsymbol{\varepsilon}_t^* = \widehat{\boldsymbol{\Gamma}}'\overline{\boldsymbol{\varepsilon}}_t$. The orthogonalized statistics $\hat{\tau}_{IV,i}$ are

$$\hat{\tau}_{IV,i} = \frac{\sum_{t=p+2}^{T} h_i\big(\widetilde{y}^{\mu}_{i,t-1}\big)\varepsilon^*_{i,t}}{\sqrt{\sum_{t=p+2}^{T} h_i^2\big(\widetilde{y}^{\mu}_{i,t-1}\big)}},$$

where $\varepsilon^*_{i,t}$ are the $N$ elements of $\boldsymbol{\varepsilon}^*_t$. According to Shin and Kang (2006), these are equivalent to using as instruments transformations of the lagged levels standardized with the residual standard deviation, i.e. $\hat{\tau}_{IV,i} = \sum_{t=p+2}^{T} h_i\big(\widetilde{y}^{\mu}_{i,t-1}/\widehat{\sigma}_{ii}\big)\varepsilon^*_{i,t}/\sqrt{\sum_{t=p+2}^{T} h_i^2\big(\widetilde{y}^{\mu}_{i,t-1}/\widehat{\sigma}_{ii}\big)}$, where $\widehat{\sigma}^2_{ii}$ is the $i$th diagonal element of $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$; define $\boldsymbol{\tau}_{IV} = (\hat{\tau}_{1,IV}, \ldots, \hat{\tau}_{N,IV})'$ as the vector stacking the individual orthogonalized statistics. Under their conditions, the asymptotic distribution of $\boldsymbol{\tau}_{IV}$ is multivariate normal with zero mean and unity covariance matrix for fixed $N$. The following panel tests studied by Shin and Kang are also available under our assumptions:

$$\overline{\tau}_{IV} = N^{-1/2}\sum_{i=1}^{N} \hat{\tau}_{IV,i},$$

as well as the Fisher-type statistic $P_{IV} = -2\sum_{i=1}^{N} \ln\big(\Phi(\hat{\tau}_{IV,i})\big)$, with $\Phi$ the standard normal cdf. We do not study their Wald-type statistic $W_{IV}$, which, being two-sided, has lower power.

Under the simplifying assumption of a fixed $N$, a panel test could be seen as rather a time series problem. While we do not share the view that such assumptions—to make asymptotics more tractable—render tests unusable, they obviously do not cover all possible $N, T$ combinations, and we now provide a joint asymptotic analysis. We require panel-specific assumptions regarding the innovations; concretely, we assume a factor structure of the panel innovations.

**Assumption 3.** *Let* $\boldsymbol{\varepsilon}_t := \boldsymbol{\Lambda}'\boldsymbol{\nu}_t + \tilde{\boldsymbol{\varepsilon}}_t$ *with* $\boldsymbol{\Lambda} = \{\boldsymbol{\lambda}_i'\}_{i=1,\ldots,N}$ *a deterministic matrix such that*

*(a)* $\boldsymbol{\lambda}_i \in \mathbb{R}^L \setminus \mathbf{0}_L$ $\forall i$, $1 \leq L$ *fixed, and* $N^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} \to \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}} > 0$;

*(b)* $\tilde{\varepsilon}_{i,t}$, $i = 1, \ldots, N$ *and* $\nu_{l,t}$, $l = 1, \ldots, L$, *are independent and they all satisfy Assumption 1.*

Requirements similar to Assumption 3(a) have been used by Bai and Ng (2004), but they require i.i.d. errors while we allow for unconditional heteroskedasticity: the innovations $\boldsymbol{\varepsilon}_t$ have at time $t$ a covariance matrix $\mathrm{E}\big(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\big) = \boldsymbol{\Sigma}(t/T)$, where $\boldsymbol{\Sigma}(t/T)$ has the typical structure of a covariance matrix in a factor model. Their "average" covariance is

$$\overline{\boldsymbol{\Omega}} = \int_0^1 \boldsymbol{\Sigma}(s)\mathrm{d}s.$$

The covariance matrix $\boldsymbol{\Sigma}(t/T)$ is time-varying; but the sample covariance approaches $\overline{\boldsymbol{\Omega}}$ in a certain sense as $N, T \to \infty$, so orthogonalization works asymptotically; see the following Lemma and the proof of Proposition 2 for details. The panel exhibits strong cross-correlation: the matrix norm $\left\|\overline{\boldsymbol{\Omega}}\right\|$ induced by the Euclidean vector norm is proportional to $N$ under Assumption 3.

**Lemma 1.** *It holds under Assumption 3 as $N, T \to \infty$ that*

$$\left\| \frac{1}{T} \sum_{t=p+2}^{T} \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' - \overline{\boldsymbol{\Omega}} \right\| = O_p\left(NT^{-0.5}\right).$$

**Proof:** See the Appendix.

The uniform higher-order cross-product moment conditions implied by independence of the idiosyncratic factors together with the summability conditions implied by Assumption 3(b) ensure the degree of homogeneity across the panel that is sufficient for joint asymptotics. In the framework of Shin and Kang (2006), fixed-$N$ asymptotics do not resort to such assumptions since $T \to \infty$ leads to joint normality, and correlation is the only form of cross-sectional dependence.

The main result of the section is given in the following Proposition about the behavior of $\overline{\tau}_{IV}$ under joint asymptotics.

**Proposition 2.** *Under Assumptions 2 and 3, it holds as $N, T \to \infty$ such that $N/T^{1/5} \to 0$ that*

$$\overline{\tau}_{IV} \xrightarrow{d} \mathcal{N}(0,1)$$

**Proof:** See the Appendix.

**Remark 1.** Assumption 3(b) could be slightly relaxed at the cost of more restrictions on the rate for for $N$: the less approximation error (cf. the proof of Proposition 2) is present in each single-unit statistic, the smaller their effect cumulated across the panel, and the more units (i.e. higher $N$-rates) can be considered without affecting $\overline{\tau}_{IV}$'s asymptotic standard normality under the null. The gain in generality is small, however, and we do not pursue this topic here.

**Remark 2.** It would alternatively be possible to construct a GLS-type panel test based on orthogonalizing the panel innovations at each time $t$ with estimates of the time-dependent covariance matrix $\mathrm{E}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)$, similar to the use of an estimated $\omega$ in the univariate case as in Boswijk (2005). The key issue in our case is to estimate $\mathrm{E}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)$ so as to preserve the mds property of the orthogonalized innovations; conveniently, Boswijk (2005) uses adaptive (recursive) estimation.

The upper bound $N = o\left(T^{1/5}\right)$ suggests that $T$ should be much larger than $N$ in small samples too; it is the consequence of having to estimate $N(N-1)/2$ covariances and computing an LU decomposition. And $N$ must be in any case smaller than $T$ to ensure positive definiteness of the sample covariance matrix. Should $N > T$, it suggests itself to make simplifying assumptions about $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ to ensure a positive definite estimate. E.g. Hartung (1999) assumes equicorrelation; his method allows to easily combine standard normal $t$-type statistics, and it is only natural to do so with the dependent single unit statistics $t^\mu_{IV,i}$. The simplification is extreme, but the method is quite robust to deviations from equicorrelation; cf. Hartung (1999) and Demetrescu et al. (2006). Alternatively, we can use shrinkage covariance estimators; see the following section.

**Remark 3.** Given the assumed rate of $N = o\left(T^{1/5}\right)$, one stationary unit with $t^\mu_{IV,i}$ diverging at rate $\sqrt{T}$ ensured by a fixed fraction of the units being under the alternative.

Panel tests have been shown to have higher power than their univariate counterparts. E.g. the first-generation test by Im et al. (2003) has power against alternatives $\rho_i = 1 - c_i/\sqrt{NT^2}$. The local power of the panel Cauchy test in general depends on nuisance parameters in the cross-dependent case due to the orthogonalization step. But it has nontrivial power in $1/\sqrt{NT^2}$ neighborhoods of the null as well, as the following proposition for the case of cross-sectional independence indicates.

**Proposition 3.** *Let $\rho_i = 1 - \frac{c_i}{\sqrt{NT^2}}$ with $0 \leq c_i \leq C\ \forall i$. Under Assumptions 2 and 3 with $\boldsymbol{\Lambda} = \mathbf{O}$, it holds as $N, T \to \infty$ such that $N/T^{1/5} \to 0$ that*

$$\overline{\tau}_{IV} \xrightarrow{d} \mathcal{N}\left(-\overline{\mu}, 1\right),$$

*where $\overline{\mu} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \frac{c_i \mu_i}{1 - \sum_{j=1}^p a_{i,j}}$, with $\mu_i = \mathrm{E}\left(\int_0^1 \mathrm{sgn}\left(\widetilde{W}^\mu_{\eta_i}(s)\right) W\left(\eta_i(s)\right) \mathrm{d}s\right)$ and $\widetilde{W}^\mu_{\eta_i} = W\left(\eta_i(s)\right) - \frac{1}{s}\int_0^s W\left(\eta_i(r)\right) \mathrm{d}r$ (the recursively demeaned time-transformed Wiener process).*

**Proof:** See the Appendix.

Thus, $\overline{\tau}_{IV}$ has good local power properties. In particular, for homoskedastic ($\eta_i(s) = s$) and homogenous ($c_i = c$) alternatives without short-run dynamics ($a_{i,j} = 0$), we obtain $\overline{\mu} = c\mu$ with $\mu = \mathrm{E}\left(\int_0^1 \mathrm{sgn}\left(\widetilde{W}^\mu(s)\right) W(s)\,\mathrm{d}s\right)$. By simulation, we find $\mu \approx 0.461$. Hence, $\overline{\tau}_{IV}$ has higher local power than the IPS test of Im et al. (2003), for which Harris et al. (2010) find $\mu \approx 0.282$ under a negligible initial condition, the most favorable case for the Im et al. (2003) test.

# 4 Small-sample behavior

Since we fit a constant throughout, we assume w.l.o.g. that $\mathrm{E}\left(y_{i,t}\right) = 0$ in our DGP:

$$y_{i,t} = \rho_i y_{i,t-1} + \varepsilon_{i,t} \qquad i = 1, \ldots, N,\ t = 1, \ldots, T$$

The variance-breaking error processes are independent normal variates $\tilde{\varepsilon}_{i,t}$, where $\mathrm{Var}(\tilde{\varepsilon}_{i,t}) = 1$ for $t = 1, \ldots, \lfloor \zeta_i T \rfloor$ and $\mathrm{Var}(\tilde{\varepsilon}_{i,t}) = 1/\delta^2$ for $t = \lfloor \zeta_i T \rfloor + 1, \ldots, T$. We consider $\delta \in \{1/5, 1, 5\}$ and take $\zeta_i = \zeta \in \{0.1, 0.5, 0.9\}$ or draw heterogeneous break dates randomly at $\zeta_i \sim U[0.1, 0.9]$. We consider two patterns of cross-sectional correlation among the $\varepsilon_{i,t}$: *A. Independence*: $\varepsilon_{i,t} = \tilde{\varepsilon}_{i,t}$. *B. Factor Structure*: $\varepsilon_{i,t} := \lambda_i \cdot \nu_t + \tilde{\varepsilon}_{i,t}$, where $\nu_t$ are i.i.d. $\mathcal{N}(0,1)$ and $\lambda_i \sim \mathcal{U}(-1,3)$. When $(\phi_1, \ldots, \phi_N)' = \mathbf{0}_N$, we study the size of the tests. To analyze power, we draw the $\phi_i$ from the uniform distribution on $[-0.1, 0]$.

Table 1: Size and Power of Second Generation Panel Tests

| | | Independence | | | | | | | | Factor Structure | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Size | | | | Power | | | | Size | | | | Power | | | |
| | $T$ \ $N$ | 6 | 16 | 26 | 46 | 6 | 16 | 26 | 46 | 6 | 16 | 26 | 46 | 6 | 16 | 26 | 46 |
| $\delta = 1/5$ | | | | | | | | | | | | | | | | | |
| | 50 | .031 | .053 | .063 | .012 | .070 | .110 | .123 | .031 | .057 | .078 | .101 | .119 | .130 | .157 | .205 | .262 |
| $S$ | 100 | .030 | .084 | .069 | .016 | .153 | .306 | .234 | .093 | .061 | .092 | .108 | .132 | .268 | .395 | .454 | .534 |
| | 200 | .028 | .087 | .062 | .015 | .451 | .629 | .546 | .401 | .058 | .111 | .109 | .125 | .667 | .880 | .916 | .971 |
| | 50 | .086 | .035 | .026 | .011 | .181 | .104 | .140 | .047 | .078 | .042 | .032 | .020 | .178 | .132 | .150 | .135 |
| $CIPS^\star$ | 100 | .078 | .026 | .026 | .010 | .394 | .372 | .542 | .603 | .069 | .024 | .027 | .018 | .379 | .335 | .529 | .603 |
| | 200 | .069 | .020 | .033 | .014 | .847 | .930 | .992 | 1.00 | .064 | .028 | .033 | .017 | .769 | .890 | .985 | .999 |
| | 50 | .044 | .077 | .113 | .030 | .102 | .187 | .195 | .076 | .024 | .024 | .006 | .002 | .127 | .184 | .207 | .249 |
| $DHT$ | 100 | .044 | .107 | .107 | .034 | .219 | .416 | .328 | .205 | .024 | .013 | .004 | .000 | .348 | .578 | .752 | .903 |
| | 200 | .038 | .123 | .098 | .034 | .579 | .748 | .724 | .658 | .026 | .014 | .005 | .001 | .785 | .980 | .996 | 1.00 |
| | 50 | .163 | .115 | .105 | .100 | .529 | .691 | .738 | .633 | .119 | .114 | .115 | .097 | .557 | .722 | .890 | .971 |
| $MPb$ | 100 | .172 | .152 | .120 | .120 | .710 | .899 | .853 | .764 | .148 | .129 | .111 | .108 | .822 | .931 | .992 | .999 |
| | 200 | .163 | .158 | .152 | .131 | .800 | .950 | .916 | .885 | .138 | .138 | .109 | .109 | .934 | .985 | .998 | 1.00 |
| $\delta = 1$ | | | | | | | | | | | | | | | | | |
| | 50 | .045 | .042 | .036 | .041 | .070 | .066 | .058 | .066 | .051 | .048 | .050 | .045 | .081 | .077 | .072 | .083 |
| $S$ | 100 | .044 | .048 | .041 | .040 | .145 | .121 | .118 | .119 | .049 | .053 | .048 | .047 | .166 | .164 | .175 | .179 |
| | 200 | .046 | .040 | .040 | .039 | .464 | .471 | .445 | .439 | .045 | .044 | .053 | .045 | .579 | .651 | .700 | .754 |
| | 50 | .104 | .054 | .080 | .069 | .185 | .130 | .189 | .188 | .107 | .066 | .082 | .076 | .214 | .189 | .251 | .266 |
| $CIPS^\star$ | 100 | .097 | .051 | .069 | .062 | .415 | .489 | .698 | .824 | .116 | .061 | .076 | .072 | .422 | .492 | .692 | .816 |
| | 200 | .099 | .047 | .068 | .061 | .881 | .985 | 1.00 | 1.00 | .102 | .053 | .074 | .065 | .808 | .955 | .992 | 1.00 |
| | 50 | .068 | .071 | .071 | .078 | .102 | .105 | .096 | .109 | .059 | .085 | .087 | .099 | .121 | .124 | .113 | .123 |
| $DHT$ | 100 | .069 | .082 | .074 | .080 | .209 | .207 | .220 | .223 | .062 | .080 | .082 | .096 | .243 | .282 | .306 | .309 |
| | 200 | .066 | .072 | .075 | .072 | .586 | .668 | .662 | .693 | .061 | .072 | .085 | .092 | .716 | .873 | .918 | .959 |
| | 50 | .113 | .069 | .055 | .060 | .572 | .706 | .771 | .843 | .091 | .055 | .053 | .049 | .619 | .870 | .934 | .980 |
| $MPb$ | 100 | .102 | .088 | .064 | .060 | .724 | .830 | .866 | .894 | .095 | .065 | .058 | .051 | .827 | .955 | .979 | .995 |
| | 200 | .110 | .089 | .080 | .071 | .845 | .922 | .936 | .955 | .098 | .074 | .057 | .067 | .923 | .982 | .996 | .997 |
| $\delta = 5$ | | | | | | | | | | | | | | | | | |
| | 50 | .052 | .080 | .047 | .186 | .070 | .093 | .072 | .174 | .188 | .193 | .270 | .375 | .138 | .139 | .165 | .196 |
| $S$ | 100 | .052 | .100 | .070 | .228 | .143 | .173 | .133 | .269 | .209 | .215 | .269 | .382 | .229 | .235 | .254 | .313 |
| | 200 | .073 | .113 | .071 | .247 | .456 | .480 | .442 | .570 | .203 | .223 | .298 | .392 | .602 | .696 | .750 | .789 |
| | 50 | .448 | .697 | .904 | .847 | .374 | .443 | .610 | .691 | .449 | .391 | .540 | .764 | .438 | .430 | .614 | .772 |
| $CIPS^\star$ | 100 | .515 | .770 | .938 | .905 | .618 | .815 | .945 | .983 | .488 | .431 | .611 | .826 | .608 | .736 | .900 | .979 |
| | 200 | .535 | .792 | .951 | .929 | .944 | .999 | 1.00 | 1.00 | .514 | .461 | .645 | .842 | .869 | .972 | .994 | 1.00 |
| | 50 | .073 | .114 | .085 | .249 | .099 | .133 | .118 | .213 | .208 | .213 | .241 | .280 | .171 | .154 | .154 | .159 |
| $DHT$ | 100 | .072 | .137 | .109 | .286 | .194 | .249 | .223 | .334 | .227 | .229 | .242 | .270 | .291 | .317 | .310 | .325 |
| | 200 | .098 | .158 | .110 | .303 | .575 | .631 | .623 | .693 | .219 | .228 | .251 | .267 | .709 | .848 | .892 | .917 |
| | 50 | .078 | .036 | .030 | .037 | .556 | .614 | .641 | .814 | .058 | .039 | .035 | .029 | .574 | .822 | .907 | .955 |
| $MPb$ | 100 | .085 | .065 | .039 | .038 | .686 | .779 | .755 | .908 | .068 | .046 | .040 | .034 | .799 | .934 | .963 | .980 |
| | 200 | .084 | .058 | .051 | .042 | .781 | .887 | .871 | .971 | .067 | .050 | .037 | .039 | .893 | .970 | .984 | .993 |

Nominal 5% level. 5000 replications. $\zeta_i \sim U[0.1, 0.9]$. $S$ is from Hanck (201x), $CIPS^\star$ is from Pesaran (2007), $DHT$ from Demetrescu et al. (2006) and $MPb$ is from Moon and Perron (2004).

Table 1 reports results for some second-generation tests (i.e. tests robust to cross-sectional dependence, but that are not designed to handle nonstationary volatility) for $\zeta_i \sim U[0.1, 0.9]$. Similar results for the other DGPs described above are available upon request. All tests handle the benchmark homoskedastic case $\delta = 1$. (For $\delta = 1$, the small-sample size distortions arise for instance because Pesaran (2007) tabulates critical values starting with $N = 10$, and we employ these for $N = 6$.) The panels for the variance breaks $\delta = 1/5$ and $\delta = 5$ however clearly demonstrate that second-generation tests do not yield valid inference under nonstationary volatility.

We therefore now turn our attention to robust tests. As regards $\overline{\tau}_{IV}$, the issue of interest is the behavior of the orthogonalization procedure, so we simulate without short-run dynamics. We nevertheless include one lagged difference to capture the effect of not knowing the true lag order in practice. Hartung's (1999) approach to capture cross-sectional dependence assumes constant correlation. He proposes to estimate the off-diagonal element $\xi$ of the correlation matrix by $\hat{\xi}^* = \max(-1/(N-1), \hat{\xi})$, where $\hat{\xi} = 1 - 1/(N-1)\sum_{i=1}^{N}\left(t^{\mu}_{IV,i} - N^{-1}\sum_{i=1}^{N} t^{\mu}_{IV,i}\right)^2$ to form the panel test statistic:

$$t_{\hat{\xi}^*,\kappa} = \frac{\sum_{i=1}^{N} t^{\mu}_{IV,i}}{\sqrt{N + (N^2 - N)\left(\hat{\xi}^* + \kappa\sqrt{\frac{2}{N+1}}(1-\hat{\xi}^*)\right)}};$$

here, $\kappa = 0.1 \cdot (1 + 1/(N+1) - \hat{\xi}^*)$ improves the small sample behavior of $t_{\hat{\xi}^*,\kappa}$. The test rejects for large negative values using standard normal critical values; see also Demetrescu et al. (2006).

Table 2 reports rejection rates for Shin and Kang's (2006) $\overline{\tau}_{IV}$, $P_{IV}$ and $t_{\hat{\xi}^*,\kappa}$ based on the $t^{\mu}_{IV,i}$. Size is well-controlled under both independence and cross-sectional dependence; $\overline{\tau}_{IV}$ is somewhat more accurate than $P_{IV}$ or $t_{\hat{\xi}^*,\kappa}$. As to power, all tests are consistent as $T \to \infty$ for any configuration of $\zeta$ and $\delta$; power increases in $N$ for $T$ sufficiently large. Once more, $\overline{\tau}_{IV}$ emerges as the most attractive choice: its power tends to be higher than that of the other tests, although there are cases where $P_{IV}$ is more powerful. The $t_{\hat{\xi}^*,\kappa}$ test seems to have lower power.

As pointed out above, the key drawback of $\overline{\tau}_{IV}$ is the requirement that $T > N$ for $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^{-1}$ to exist. This may not be the case in practice. Moreover, if $T$ is only moderately larger than $N$, the finite-sample performance of $\overline{\tau}_{IV}$ will suffer. We therefore employ a recent proposal by Ledoit and Wolf (2004) to estimate $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$ allowing in principle any configuration of $T$ and $N$. They propose to construct a weighted version of $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$ and the identity matrix $\boldsymbol{I}$, $\boldsymbol{S}_T = \kappa_{1T}\boldsymbol{I} + \kappa_{2T}\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$. Specifically, $\kappa_{1T}$ and $\kappa_{2T}$ are constructed as follows. Define

$$\bar{b}_T^2 = \frac{1}{N}\left[\sum_{t=p+2}^{T}\left(\frac{\overline{\boldsymbol{\varepsilon}}_t'\overline{\boldsymbol{\varepsilon}}_t}{T}\right)^2 - \frac{1}{T}\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^2)\right].$$

Further, $m_T = \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_{\varepsilon})/N$, $d_T^2 = \operatorname{tr}[(\widehat{\boldsymbol{\Sigma}}_{\varepsilon} - m_T\boldsymbol{I})(\widehat{\boldsymbol{\Sigma}}_{\varepsilon} - m_T\boldsymbol{I})']/N$, $b_T^2 = \min(\bar{b}_T^2, d_T^2)$ and $a_T^2 = d_T^2 - b_T^2$. Then, $\kappa_{1T} = m_T \cdot b_T^2/d_T^2$ and $\kappa_{2T} = a_T^2/d_T^2$. The full-rank matrix $\boldsymbol{I}$ ensures that $\boldsymbol{S}_T$ is invertible even if $T < N$. The (generally misspecified, but invertible) structure imposed by adding

Table 2: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests

| | | Independence | | | | | | | | Factor Structure | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Size | | | | Power | | | | Size | | | | Power | | | |
| | $T$ $N$ | 6 | 16 | 26 | 46 | 6 | 16 | 26 | 46 | 6 | 16 | 26 | 46 | 6 | 16 | 26 | 46 |
| $\delta = 1/5$ | | | | | | | | | | | | | | | | | |
| | 50 | .054 | .049 | .039 | .044 | .175 | .244 | .235 | .159 | .043 | .042 | .038 | .029 | .268 | .400 | .454 | .387 |
| $\overline{\tau}_{IV}$ | 100 | .053 | .059 | .055 | .045 | .359 | .526 | .510 | .456 | .047 | .040 | .040 | .034 | .608 | .840 | .951 | .984 |
| | 200 | .049 | .058 | .052 | .052 | .593 | .852 | .840 | .843 | .048 | .047 | .049 | .040 | .922 | .996 | 1.00 | 1.00 |
| | 50 | .047 | .047 | .047 | .041 | .172 | .228 | .236 | .170 | .037 | .038 | .032 | .035 | .195 | .287 | .347 | .342 |
| $P_{IV}$ | 100 | .049 | .057 | .051 | .043 | .392 | .534 | .600 | .648 | .042 | .037 | .039 | .032 | .496 | .746 | .891 | .953 |
| | 200 | .049 | .052 | .053 | .048 | .701 | .900 | .935 | .971 | .045 | .042 | .043 | .041 | .882 | .992 | 1.00 | 1.00 |
| | 50 | .049 | .059 | .054 | .041 | .105 | .106 | .099 | .113 | .040 | .035 | .029 | .023 | .133 | .135 | .135 | .110 |
| $t_{\hat{\xi}^*,\kappa}$ | 100 | .055 | .066 | .062 | .048 | .255 | .265 | .237 | .287 | .047 | .043 | .037 | .025 | .290 | .336 | .359 | .379 |
| | 200 | .061 | .072 | .054 | .049 | .574 | .606 | .619 | .690 | .045 | .043 | .037 | .028 | .659 | .807 | .892 | .964 |
| $\delta = 1$ | | | | | | | | | | | | | | | | | |
| | 50 | .046 | .046 | .045 | .048 | .177 | .219 | .243 | .214 | .041 | .034 | .035 | .027 | .268 | .417 | .489 | .439 |
| $\overline{\tau}_{IV}$ | 100 | .048 | .050 | .048 | .051 | .363 | .483 | .566 | .664 | .046 | .043 | .036 | .034 | .566 | .850 | .932 | .978 |
| | 200 | .048 | .047 | .050 | .054 | .636 | .830 | .905 | .969 | .051 | .052 | .044 | .042 | .897 | .992 | .999 | 1.00 |
| | 50 | .041 | .038 | .041 | .043 | .169 | .225 | .248 | .219 | .034 | .031 | .031 | .030 | .211 | .324 | .393 | .396 |
| $P_{IV}$ | 100 | .043 | .044 | .044 | .042 | .395 | .581 | .672 | .758 | .042 | .039 | .037 | .032 | .498 | .790 | .906 | .966 |
| | 200 | .049 | .048 | .047 | .046 | .754 | .942 | .979 | .998 | .047 | .046 | .040 | .039 | .883 | .992 | 1.00 | 1.00 |
| | 50 | .046 | .038 | .038 | .035 | .120 | .111 | .110 | .109 | .042 | .049 | .045 | .041 | .142 | .126 | .124 | .098 |
| $t_{\hat{\xi}^*,\kappa}$ | 100 | .053 | .051 | .049 | .046 | .271 | .301 | .305 | .319 | .051 | .057 | .060 | .055 | .318 | .378 | .409 | .414 |
| | 200 | .057 | .059 | .054 | .050 | .624 | .725 | .717 | .760 | .055 | .054 | .063 | .065 | .726 | .877 | .931 | .967 |
| $\delta = 5$ | | | | | | | | | | | | | | | | | |
| | 50 | .045 | .041 | .039 | .040 | .173 | .176 | .157 | .192 | .040 | .033 | .029 | .029 | .192 | .315 | .362 | .311 |
| $\overline{\tau}_{IV}$ | 100 | .049 | .045 | .048 | .041 | .338 | .353 | .369 | .553 | .048 | .037 | .038 | .032 | .411 | .678 | .784 | .852 |
| | 200 | .048 | .049 | .049 | .043 | .507 | .630 | .716 | .883 | .046 | .046 | .043 | .041 | .714 | .949 | .985 | .996 |
| | 50 | .038 | .032 | .030 | .038 | .197 | .246 | .247 | .198 | .034 | .028 | .028 | .028 | .154 | .258 | .315 | .283 |
| $P_{IV}$ | 100 | .047 | .037 | .035 | .039 | .444 | .567 | .676 | .656 | .040 | .032 | .030 | .029 | .384 | .673 | .795 | .876 |
| | 200 | .047 | .044 | .040 | .038 | .697 | .877 | .964 | .962 | .044 | .043 | .039 | .036 | .734 | .971 | .995 | 1.00 |
| | 50 | .046 | .037 | .034 | .021 | .106 | .096 | .114 | .071 | .046 | .046 | .042 | .047 | .108 | .101 | .091 | .087 |
| $t_{\hat{\xi}^*,\kappa}$ | 100 | .047 | .050 | .047 | .031 | .258 | .226 | .260 | .179 | .054 | .064 | .060 | .063 | .266 | .304 | .310 | .320 |
| | 200 | .058 | .057 | .053 | .049 | .555 | .531 | .598 | .493 | .062 | .077 | .077 | .080 | .610 | .768 | .832 | .859 |

Nominal 5% level. 5000 replications. $\zeta_i \sim U[0.1, 0.9]$.

$\kappa_{1T}\boldsymbol{I}$ to the unbiased estimator $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ introduces a finite-sample bias in $\boldsymbol{S}_T$. Yet, the weights $\kappa_{1T}$ and $\kappa_{2T}$ are optimal in the sense that $\boldsymbol{S}_T$ asymptotically (for $N, T \to \infty$ jointly) has minimum expected loss in a class of linear combinations of $\boldsymbol{I}$ and $\widehat{\boldsymbol{\Sigma}}_\varepsilon$. Ledoit and Wolf (2004) show the joint asymptotics to be a good guide in finite samples, including the case $T < N$. Moreover, the following lemma shows $\boldsymbol{S}_T$ to converge to $\overline{\boldsymbol{\Omega}}$ at the same rate as $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ under the assumptions of Proposition 2, so it can be safely used for the test of Shin and Kang (2006).

**Lemma 2.** *Under the assumptions of Proposition 2, it holds that*

$$\left\| \boldsymbol{S}_T - \overline{\boldsymbol{\Omega}} \right\| = O_p\left(NT^{-0.5}\right).$$

**Proof:** See the Appendix.

We now present additional simulations gauging the effectiveness of Shin and Kang's (2006) tests using shrinkage, allowing us to also consider the case $T < N$. Table 3 reports rejection

Table 3: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests with shrinkage

| | | Independence | | | | | | | | Factor Structure | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Size | | | | Power | | | | Size | | | | Power | | | |
| | $T$ $N$ | 16 | 26 | 56 | 106 | 16 | 26 | 56 | 106 | 16 | 26 | 56 | 106 | 16 | 26 | 56 | 106 |
| $\delta = 1/5$ | | | | | | | | | | | | | | | | | |
| | 50 | .024 | .035 | .007 | .020 | .144 | .217 | .090 | .253 | .034 | .036 | .023 | .014 | .422 | .682 | .820 | .930 |
| $\overline{\tau}_{IV}$ | 100 | .033 | .041 | .019 | .032 | .419 | .588 | .472 | .813 | .039 | .040 | .036 | .024 | .888 | .987 | 1.00 | 1.00 |
| | 200 | .039 | .048 | .038 | .044 | .746 | .913 | .882 | .998 | .043 | .047 | .038 | .028 | .999 | 1.00 | 1.00 | 1.00 |
| | 50 | .006 | .006 | .000 | .000 | .057 | .058 | .002 | .001 | .015 | .013 | .001 | .000 | .201 | .367 | .301 | .195 |
| $P_{IV}$ | 100 | .014 | .013 | .001 | .000 | .431 | .458 | .441 | .328 | .018 | .022 | .005 | .000 | .730 | .929 | .979 | .996 |
| | 200 | .029 | .028 | .014 | .006 | .873 | .946 | .977 | .997 | .030 | .031 | .011 | .003 | .995 | 1.00 | 1.00 | 1.00 |
| | 50 | .045 | .049 | .040 | .048 | .106 | .099 | .114 | .092 | .032 | .028 | .020 | .020 | .143 | .134 | .092 | .050 |
| $t_{\hat{\xi}^*,\kappa}$ | 100 | .053 | .052 | .044 | .052 | .254 | .251 | .285 | .248 | .039 | .032 | .028 | .024 | .351 | .363 | .335 | .364 |
| | 200 | .053 | .069 | .055 | .063 | .643 | .630 | .691 | .666 | .043 | .035 | .029 | .029 | .803 | .910 | .962 | .997 |
| $\delta = 1$ | | | | | | | | | | | | | | | | | |
| | 50 | .034 | .030 | .016 | .005 | .190 | .215 | .198 | .137 | .029 | .024 | .009 | .001 | .405 | .484 | .551 | .463 |
| $\overline{\tau}_{IV}$ | 100 | .039 | .037 | .034 | .026 | .488 | .562 | .672 | .731 | .041 | .032 | .027 | .014 | .843 | .933 | .986 | .996 |
| | 200 | .041 | .048 | .048 | .046 | .817 | .892 | .973 | .995 | .044 | .044 | .041 | .039 | .993 | .999 | 1.00 | 1.00 |
| | 50 | .013 | .007 | .000 | .000 | .124 | .115 | .032 | .000 | .013 | .008 | .000 | .000 | .238 | .259 | .138 | .002 |
| $P_{IV}$ | 100 | .025 | .020 | .010 | .001 | .536 | .623 | .658 | .459 | .031 | .022 | .011 | .001 | .760 | .877 | .959 | .950 |
| | 200 | .034 | .036 | .029 | .018 | .929 | .974 | .997 | 1.00 | .038 | .036 | .032 | .019 | .996 | .999 | 1.00 | 1.00 |
| | 50 | .043 | .039 | .034 | .027 | .106 | .108 | .107 | .103 | .040 | .042 | .044 | .034 | .131 | .109 | .082 | .059 |
| $t_{\hat{\xi}^*,\kappa}$ | 100 | .051 | .045 | .040 | .045 | .314 | .305 | .318 | .321 | .047 | .056 | .059 | .057 | .391 | .404 | .403 | .424 |
| | 200 | .055 | .057 | .047 | .051 | .702 | .718 | .763 | .775 | .061 | .064 | .073 | .073 | .879 | .925 | .975 | .987 |
| $\delta = 5$ | | | | | | | | | | | | | | | | | |
| | 50 | .014 | .017 | .009 | .001 | .134 | .168 | .193 | .092 | .025 | .020 | .005 | .001 | .275 | .319 | .344 | .261 |
| $\overline{\tau}_{IV}$ | 100 | .026 | .026 | .028 | .007 | .359 | .442 | .595 | .588 | .033 | .032 | .019 | .009 | .627 | .743 | .890 | .929 |
| | 200 | .040 | .041 | .034 | .022 | .614 | .784 | .931 | .957 | .038 | .037 | .032 | .025 | .933 | .980 | .998 | 1.00 |
| | 50 | .003 | .002 | .000 | .000 | .073 | .054 | .017 | .000 | .010 | .004 | .000 | .000 | .146 | .131 | .038 | .000 |
| $P_{IV}$ | 100 | .010 | .006 | .003 | .000 | .469 | .446 | .432 | .269 | .022 | .015 | .004 | .000 | .578 | .704 | .808 | .676 |
| | 200 | .023 | .019 | .013 | .001 | .833 | .908 | .970 | .979 | .029 | .027 | .019 | .006 | .955 | .991 | 1.00 | 1.00 |
| | 50 | .042 | .028 | .019 | .035 | .100 | .086 | .053 | .106 | .048 | .048 | .047 | .039 | .098 | .097 | .077 | .077 |
| $t_{\hat{\xi}^*,\kappa}$ | 100 | .047 | .042 | .032 | .038 | .249 | .211 | .189 | .280 | .068 | .070 | .067 | .067 | .286 | .322 | .312 | .305 |
| | 200 | .059 | .050 | .042 | .050 | .589 | .533 | .521 | .665 | .066 | .074 | .082 | .080 | .752 | .805 | .858 | .888 |

Nominal 5% level. 5000 replications. $\zeta_i \sim U[0.1, 0.9]$.

rates for $N \in \{16, 26, 56, 106\}$. The $P_{IV}$ test is now sometimes drastically undersized especially for $N \gg T$. Reassuringly, this does not destroy its consistency as $P_{IV}$ remains powerful at least for large $T$. On the other hand, $\overline{\tau}_{IV}$ mostly performs quite well even with shrinkage and in cases where $N > T$, although predictably somewhat less accurately than when one can use an estimator $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ that unbiasedly estimates the true covariance matrix. In terms of size, $t_{\hat{\xi}^*,\kappa}$ not requiring shrinkage emerges as a serious competitor when $N > T$. However, $\overline{\tau}_{IV}$ is substantially more powerful than $t_{\hat{\xi}^*,\kappa}$ for small and intermediate $T$ whenever size is comparable. Overall, these results lead us to recommend to employ $\overline{\tau}_{IV}$ in cross-dependent panels.

## 5 Concluding remarks

The Cauchy estimator, for which the sign of the lagged level instruments the lagged level itself, yields a unit root test with an asymptotic standard normal null distribution even under unconditional heteroskedasticity.

The paper showed that the features of the Cauchy test extend in cross-dependent, heteroskedastic panels. In particular, we prove the panel unit root test due to Shin and Kang (2006) to be robust to unconditional heteroskedasticity. Moreover, the test was shown to be locally more powerful than the IPS test of Im et al. (2003).

The assumptions under which joint $N, T$ asymptotics hold suggested that $N$ should be smaller than $T$. To extend the applicability of the panel test to situations where $T$ is comparable to, or smaller than, $N$, we proposed the use of shrinkage covariance matrix estimators. The test performed well in small samples.

## Proofs

Note: Sums run from $t = p + 2$ to $T$ unless specified otherwise, and $C$ stands for a generic constant.

### Proof of Lemma 1

Note that it suffices to show that $T^{-1} \sum \varepsilon_{i,t}\varepsilon_{j,t}$ is $\sqrt{T}$-consistent at a uniform rate over $1 \leq i, j \leq N$ (recall that the norm of an $N \times N$ matrix with uniformly bounded elements is $O(N)$). To this end, we make use of the factor structure of the innovations. We namely have that

$$T^{-1} \sum \varepsilon_{i,t}\varepsilon_{j,t} = T^{-1} \sum \boldsymbol{\lambda}_i' \boldsymbol{\nu}_t \boldsymbol{\nu}_t' \boldsymbol{\lambda}_j + T^{-1} \sum \boldsymbol{\lambda}_i' \boldsymbol{\nu}_t \widetilde{\varepsilon}_{j,t} + T^{-1} \sum \widetilde{\varepsilon}_{i,t} \boldsymbol{\lambda}_j' \boldsymbol{\nu}_t + T^{-1} \sum \widetilde{\varepsilon}_{i,t}\widetilde{\varepsilon}_{j,t}.$$

Now,

$$\begin{aligned}
\mathrm{Var}\left(T^{-1} \sum \widetilde{\varepsilon}_{i,t}^2 - \overline{\omega}_i^2\right) &= \mathrm{Var}\left(T^{-1} \sum \widetilde{\varepsilon}_{i,t}^2 - T^{-1} \sum \omega_i^2(t/T)\right) \\
&= T^{-2} \sum_{t=p+2}^{T} \sum_{s=p+2}^{T} \omega_i^2(t/T)\,\omega_i^2(s/T)\,\mathrm{E}\left(\left(\widetilde{\epsilon}_{i,t}^2 - 1\right)\left(\widetilde{\epsilon}_{i,s}^2 - 1\right)\right);
\end{aligned}$$

using the obvious uniform boundedness of $\omega_i$ across the panel, we obtain that

$$\mathrm{Var}\left(\frac{1}{T} \sum \widetilde{\varepsilon}_{i,t}^2 - \overline{\omega}_i^2\right) \leq \frac{C}{T^2} \sum_{t=p+2}^{T} \sum_{s=p+2}^{T} \mathrm{E}\left(\left|\widetilde{\epsilon}_{i,t}^2 - 1\right| \left|\widetilde{\epsilon}_{i,s}^2 - 1\right|\right)$$

irrespective of $i$. The same reasoning applies to the other cross-products as well, leading with the summability conditions in Assumption 3 (b) to $\sup_{1 \leq i,j \leq N} \mathrm{Var}\left(T^{-1} \sum \varepsilon_{i,t}\varepsilon_{j,t} - \overline{\boldsymbol{\Omega}}_{i,j}\right) \leq C/T$. Thus the sample covariances of $\boldsymbol{\varepsilon}_t$ are $\sqrt{T}$-consistent for the respective elements of $\overline{\boldsymbol{\Omega}}$; as required.

### Proof of Proposition 2

Let us first analyze the behavior of the sample covariance matrix of $\overline{\boldsymbol{\varepsilon}}_t$, $\widehat{\boldsymbol{\Sigma}}_{\varepsilon} = T^{-1} \sum \overline{\boldsymbol{\varepsilon}}_t \overline{\boldsymbol{\varepsilon}}_t'$. Note that $\overline{\varepsilon}_{i,t} = \varepsilon_{i,t} + O_p\left(T^{-0.5}\right)$ (whether estimating by imposing the unit root or not); we have at the assumed maximal rate for $N$ that $\left\|\frac{1}{T}\sum\overline{\boldsymbol{\varepsilon}}_t\overline{\boldsymbol{\varepsilon}}_t' - \frac{1}{T}\sum\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right\| = O_p\left(NT^{-0.5}\right)$, so, considering Lemma 1, it follows that

$$\left\|\widehat{\boldsymbol{\Sigma}}_\varepsilon - \overline{\boldsymbol{\Omega}}\right\| = O_p\left(NT^{-0.5}\right). \tag{4}$$

Making now use of Equation (11b) from Lütkepohl (1996, p. 107), we have that

$$\left\|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{-1} - \overline{\boldsymbol{\Omega}}^{-1}\right\| \leq \frac{\left\|\widehat{\boldsymbol{\Sigma}}_\varepsilon - \overline{\boldsymbol{\Omega}}\right\|\left\|\overline{\boldsymbol{\Omega}}^{-1}\right\|^2}{1-\left\|\overline{\boldsymbol{\Omega}}^{-1}\left(\widehat{\boldsymbol{\Sigma}}_\varepsilon - \overline{\boldsymbol{\Omega}}\right)\right\|}.$$

Due to the factor structure of the innovations, $\overline{\boldsymbol{\Omega}}$ has eigenvalues bounded away from 0, hence $\left\|\overline{\boldsymbol{\Omega}}^{-1}\right\| < C$; considering that $\left\|\widehat{\boldsymbol{\Sigma}}_\varepsilon - \overline{\boldsymbol{\Omega}}\right\| \xrightarrow{p} 0$, the denominator of the r.h.s. converges in probability to 1 and the numerator to 0 at rate $O_p(NT^{-0.5})$. This implies the same convergence rate of each of the elements of $\widehat{\boldsymbol{\Sigma}}_\varepsilon^{-1}$ and of $\widehat{\boldsymbol{\Gamma}}'$. Denote by $\boldsymbol{\Gamma}'$ the $LU$ decomposition of $\overline{\boldsymbol{\Omega}}^{-1}$ and note that $\|\boldsymbol{\Gamma}\| < C$ as well. Since the elements of $\widehat{\boldsymbol{\Gamma}}'$ are consistent at rate $O_p(NT^{-0.5})$, it follows that $\left\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right\| = O_p(N^2T^{-0.5})$. Moving on to the main part of the proof, we have with $h_i\left(\widetilde{y}_{i,t-1}^\mu\right) = h_i\left(\widetilde{x}_{i,t-1}^\mu\right)$ that

$$\overline{\tau}_{IV} = \frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{\sum h_i\left(\widetilde{x}_{i,t-1}^\mu\right)\varepsilon_{i,t}^*}{\sqrt{\sum h_i^2\left(\widetilde{x}_{i,t-1}^\mu\right)}}.$$

With Lemma A.1B from Demetrescu and Hanck (2011) [DH] we have that $h_i^2\left(\widetilde{x}_{i,t-1}^\mu\right) = 1 + O_p(t^{-0.5})$ and hence $\sum h_i^2\left(\widetilde{x}_{i,t-1}^\mu\right) = T + O_p(T^{0.5})$. Since, as can easily be deduced using the same arguments as for the derivations below, $\sum h_i\left(\widetilde{x}_{i,t-1}^\mu\right)\varepsilon_{i,t}^* = O_p(\sqrt{T})$, it follows from dividing denominator and numerator by $1/\sqrt{T}$ and using a Taylor expansion that

$$\frac{\sum h_i\left(\widetilde{x}_{i,t-1}^\mu\right)\varepsilon_{i,t}^*}{\sqrt{\sum h_i^2\left(\widetilde{x}_{i,t-1}^\mu\right)}} = \frac{1}{\sqrt{T}}\sum h_i\left(\widetilde{x}_{i,t-1}^\mu\right)\varepsilon_{i,t}^* + O_p(T^{-0.5});$$

furthermore, with obvious notation $\mathbf{h}_t = \left(h_i(\widetilde{x}_{i,t-1}^\mu)\right)_{i=1,\ldots,N}'$,

$$\begin{aligned}\overline{\tau}_{IV} &= \frac{1}{\sqrt{TN}}\sum_{i=1}^N\sum h_i\left(\widetilde{x}_{i,t-1}^\mu\right)\varepsilon_{i,t}^* + O_p\left(T^{-0.5}N^{0.5}\right)\\ &= \frac{1}{\sqrt{TN}}\sum_{i=1}^N\sum h_i\left(\widetilde{x}_{i,t-1}^\mu\right)\varepsilon_{i,t}^* + o_p(1) = \frac{1}{\sqrt{TN}}\sum \mathbf{h}_t'\widehat{\boldsymbol{\Gamma}}'\overline{\boldsymbol{\varepsilon}}_t + o_p(1).\end{aligned}$$

We further have that $\frac{1}{\sqrt{TN}}\sum\mathbf{h}_t'\widehat{\boldsymbol{\Gamma}}'\overline{\boldsymbol{\varepsilon}}_t = \frac{1}{\sqrt{TN}}\sum\mathbf{h}_t'\boldsymbol{\Gamma}'\overline{\boldsymbol{\varepsilon}}_t + \frac{1}{\sqrt{TN}}\sum\mathbf{h}_t'(\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}')\overline{\boldsymbol{\varepsilon}}_t$. Now,

$$\frac{1}{\sqrt{TN}}\sum\mathbf{h}_t'(\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}')\overline{\boldsymbol{\varepsilon}}_t = \operatorname{tr}\left(\frac{1}{\sqrt{N}}\left(\frac{1}{\sqrt{T}}\sum\overline{\boldsymbol{\varepsilon}}_t\mathbf{h}_t'\right)\left(\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}'\right)\right);$$

using arguments analog to those used for Lemma A.1E in DH we conclude that for all $1 \leq k, i \leq N$,

$$\begin{aligned}\sum h_k\left(\widetilde{x}_{k,t-1}^\mu\right)\overline{\varepsilon}_{i,t} &= \sum h_k\left(\widetilde{x}_{k,t-1}^\mu\right)\left(\Delta y_{i,t} - \sum\nolimits_j \overline{a}_{ij}\Delta y_{i,t-j}\right)\\ &= \sum h_k\left(\widetilde{x}_{k,t-1}^\mu\right)\left(\sum\nolimits_j (a_{ij} - \overline{a}_{ij})\Delta x_{i,t-j} + \varepsilon_{i,t}\right)\\ &= \sum h_k\left(\widetilde{x}_{k,t-1}^\mu\right)\varepsilon_{i,t} + o_p(T^{0.25}).\end{aligned}$$

So the elements of the $N \times N$ matrix $T^{-1/2}\sum\overline{\boldsymbol{\varepsilon}}_t\mathbf{h}_t'$ are uniformly bounded in probability, and

$$\left\|N^{-1/2}\left(T^{-1/2}\sum\overline{\boldsymbol{\varepsilon}}_t\mathbf{h}_t'\right)\left(\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}'\right)\right\| \leq N^{-1/2}\left\|T^{-1/2}\sum\overline{\boldsymbol{\varepsilon}}_t\mathbf{h}_t'\right\|\left\|\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}'\right\| = O_p(N^{2.5}T^{-0.5}).$$

The norm on the l.h.s. thus vanishes, implying that the trace vanishes too. Summing up, we have that

$$\overline{\tau}_{IV} = \frac{1}{\sqrt{TN}} \sum \mathbf{h}_t' \boldsymbol{\Gamma}' \overline{\boldsymbol{\varepsilon}}_t + o_p(1).$$

Using again Lemma A.1E in DH as above, it follows immediately that $\overline{\tau}_{IV} = \frac{1}{\sqrt{TN}} \sum \mathbf{h}_t' \boldsymbol{\Gamma}' \boldsymbol{\varepsilon}_t + o_p(1)$; letting $\boldsymbol{\imath}_t = \left(\mathrm{sgn}_i\left(\widetilde{x}^{\mu}_{i,t-1}\right)\right)'_{i=1,\ldots,N}$ and employing Lemma A.1C in DH, we have furthermore that

$$\overline{\tau}_{IV} = \frac{1}{\sqrt{TN}} \sum \boldsymbol{\imath}_t' \boldsymbol{\Gamma}' \boldsymbol{\varepsilon}_t + o_p(1).$$

$N^{-0.5}\boldsymbol{\imath}_t' \boldsymbol{\Gamma}' \boldsymbol{\varepsilon}_t$ is a md array; given the finiteness of its 4th order moments (cf. Assumption 3), the second condition of the CLT for md arrays (Davidson, 1994, Thm. 24.3) is fulfilled; checking the first condition amounts to showing that $T^{-1}\sum\left(N^{-0.5}\boldsymbol{\imath}_t' \boldsymbol{\Gamma}' \boldsymbol{\varepsilon}_t\right)^2 \xrightarrow{p} 1$: write the quantity as $\frac{1}{NT}\sum \boldsymbol{\imath}_t' \boldsymbol{\Gamma}' \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \boldsymbol{\Gamma} \boldsymbol{\imath}_t$ and recall that $\left\|T^{-1}\sum \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' - \left(\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\right)^{-1}\right\| = O_p\left(T^{-0.5}N\right)$, and the result follows.

## Proof of Proposition 3

Begin by examining, like in the proof of Proposition 2, the quantity

$$\frac{1}{\sqrt{NT}} \sum \mathbf{h}_t' \widehat{\boldsymbol{\Gamma}}' \overline{\boldsymbol{\varepsilon}}_t = \frac{1}{\sqrt{NT}} \sum \mathbf{h}_t' \boldsymbol{\Gamma}' \overline{\boldsymbol{\varepsilon}}_t + \mathrm{tr}\left(\frac{1}{\sqrt{N}}\left(\frac{1}{\sqrt{T}}\sum \overline{\boldsymbol{\varepsilon}}_t \mathbf{h}_t'\right)\left(\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}'\right)\right).$$

We show that the trace vanishes under the local alternative as well. Namely,

$$\begin{aligned}\sum h_k\left(\widetilde{x}^{\mu}_{k,t-1}\right)\overline{\varepsilon}_{i,t} \quad &= \quad \sum h_k\left(\widetilde{x}^{\mu}_{k,t-1}\right)\left(\varepsilon_{i,t} - \frac{c_i}{\sqrt{NT^2}} x_{i,t-1}\right) \\ &\quad + \sum_{j=1}^{p}\left(a_{ij} - \overline{a}_{ij}\right)\sum h_k\left(\widetilde{x}^{\mu}_{k,t-1}\right)\Delta x_{i,t-j}\end{aligned}$$

for all $1 \leq i, k \leq N$. With $x_{i,t-1}/\sqrt{T}$ uniformly $L_2$-bounded, we have as in Lemma A.1E in DH that

$$\sum h_k\left(\widetilde{x}^{\mu}_{k,t-1}\right)\overline{\varepsilon}_{i,t} = \sum h_k\left(\widetilde{x}^{\mu}_{k,t-1}\right)\left(\varepsilon_{i,t} - \frac{c_i}{\sqrt{NT^2}} x_{i,t-1}\right) + o_p\left(T^{0.25}\right) = O_p\left(\sqrt{T}\right).$$

Note that $\overline{\varepsilon}_{i,t} = \varepsilon_{i,t} + O_p\left(T^{-0.5}\right)$ holds under the considered local alternative too, so the proof of Lemma 1 still applies leading to $\left\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right\| = O_p\left(N^2 T^{-0.5}\right)$, and the trace does indeed vanish like in the proof of Proposition 2. Use now the Cauchy-Schwarz inequality and Lemma A.1B in DH together with the uniform $L_2$-boundedness of $x_{k,t-1}/\sqrt{T}$ (in $t$ and $i$) to establish that

$$\mathrm{E}\left|\frac{1}{\sqrt{N^2T^3}}\sum_{i=1}^{N}\sum\left(h_k\left(\widetilde{x}^{\mu}_{k,t-1}\right) - \mathrm{sgn}\left(\widetilde{x}^{\mu}_{k,t-1}\right)\right)x_{i,t-1}\right| = o(1);$$

letting $\boldsymbol{C} = \mathrm{diag}\left(c_i\right)$ (the diagonal matrix with diagonal elements $c_i$) and recalling that $\boldsymbol{\Gamma} = \mathrm{diag}\left(\overline{\omega}_i^{-1}\right)$, we have as in the proof of Proposition 2 that

$$\begin{aligned}\overline{\tau}_{IV} \quad &= \quad \frac{1}{\sqrt{NT}}\sum \mathbf{h}_t' \boldsymbol{\Gamma}'\left(\boldsymbol{\varepsilon}_t - \frac{1}{\sqrt{NT^2}}\boldsymbol{C}\mathbf{x}_{t-1}\right) + o_p(1) \\ &= \quad \frac{1}{\sqrt{NT}}\sum \boldsymbol{\imath}_t' \boldsymbol{\Gamma}' \boldsymbol{\varepsilon}_t - \frac{1}{\sqrt{N^2T^3}}\sum_{i=1}^{N} c_i \sum \mathrm{sgn}\left(\widetilde{x}^{\mu}_{i,t-1}\right)\frac{x_{i,t-1}}{\overline{\omega}_i} + o_p(1).\end{aligned}$$

The proof of Proposition 2 shows that $\frac{1}{\sqrt{NT}}\sum \boldsymbol{\imath}_t' \boldsymbol{\Gamma}' \boldsymbol{\varepsilon}_t \sim \mathcal{N}(0,1)$ as $N, T \to \infty$ with $N = o(T^{0.2})$, so we only need to examine the noncentrality term of $\overline{\tau}_{IV}$'s asymptotic distribution: use the independence of the units and the uniform $L_2$-boundedness (in $i$) of $\frac{1}{T}\sum \mathrm{sgn}\left(\widetilde{x}^{\mu}_{i,t-1}\right)\frac{x_{i,t-1}}{\sqrt{T}}$ to conclude that

$$\frac{1}{NT}\sum_{i=1}^{N} c_i \sum \mathrm{sgn}\left(\widetilde{x}^{\mu}_{i,t-1}\right)\frac{x_{i,t-1}}{\overline{\omega}_i\sqrt{T}} = \frac{1}{N}\sum_{i=1}^{N}\frac{c_i}{\overline{\omega}_i}\,\mathrm{E}\left(\frac{1}{T}\sum \mathrm{sgn}\left(\widetilde{x}^{\mu}_{i,t-1}\right)\frac{x_{i,t-1}}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right),$$

so the result follows if $\mathrm{E}\left(\frac{1}{T^{1.5}}\sum \operatorname{sgn}\left(\widetilde{x}^{\mu}_{i,t-1}\right) x_{i,t-1}\right) \to \frac{\overline{\omega}_i}{A_i}\mathrm{E}\left(\int_0^1 \operatorname{sgn}\left(\widetilde{W}^{\mu}_{\eta_i}(s)\right) W(\eta_i(s))\,\mathrm{d}s\right)$, where $A_i = 1 - \sum_{j=1}^{p} a_{i,j}$. The Beveridge-Nelson decomposition and the uniform boundedness of the variances imply

$$\mathrm{E}\left|x_{i,t} - \frac{1}{A_i}\sum\nolimits_{j=0}^{t} \rho_i^j \varepsilon_{i,t-j}\right|^2 \le C. \tag{5}$$

A Taylor series expansion for $r^{\alpha}$ with rest term in differential form, $r^{\alpha} = r_0^{\alpha} + \alpha\varrho^{\alpha-1}(r - r_0)$ with $\varrho$ between $r$ and $r_0$, gives $\rho_i^j = 1 + j\varrho_i^{j-1}(\rho_i - 1)$ for $\rho_i \le \varrho_i \le 1$. Recall, $\rho_i = 1 - \frac{c_i}{\sqrt{NT^2}}$, so $\exists N_0, T_0$ fixed such that $0 < 1 - \frac{c_i}{\sqrt{NT^2}} \le \varrho_i \le 1$ for all $N > N_0$ and $T > T_0$; thus, $0 < \varrho_i^{j-1} < 1\ \forall j$. Hence,

$$\left|\rho_i^j - 1\right| \le \frac{C}{\sqrt{NT^2}} j \qquad \text{for } N \text{ and } T \text{ large enough,}$$

leading with Minkowski's inequality and (5) to

$$\begin{aligned}
\sqrt{\mathrm{E}\left|x_{i,t} - \frac{1}{A_i}\sum\nolimits_{j=0}^{t} \varepsilon_{i,t-j}\right|^2} &\le \sqrt{\mathrm{E}\left|x_{i,t} - \frac{1}{A_i}\sum\nolimits_{j=0}^{t} \rho_i^j\varepsilon_{i,t-j}\right|^2} + C\sqrt{\mathrm{E}\left|\sum\nolimits_{j=0}^{t}\left(\rho_i^j - 1\right)\varepsilon_{i,t-j}\right|^2} \\
&\le \frac{C}{\sqrt{NT^2}}\sqrt{\sum\nolimits_{j=0}^{t} j^2 \operatorname{Var}(\varepsilon_{i,t-j})} = C\sqrt{\frac{T}{N}}
\end{aligned}$$

uniformly in $i$ and $t$, or to $\frac{1}{\sqrt{T}}x_{i,t} = \frac{1}{A_i\sqrt{T}}\sum_{j=0}^{t}\varepsilon_{i,t-j} + O_p\left(N^{-0.5}\right)$. Then, $\frac{1}{\sqrt{T}}x_{i,t} \Rightarrow \frac{\overline{\omega}_i}{A_i}W(\eta_i(s))$, implying

$$\frac{1}{T^{1.5}}\sum \operatorname{sgn}\left(\widetilde{x}^{\mu}_{i,t-1}\right)x_{i,t-1} \Rightarrow \frac{\overline{\omega}_i}{1 - \sum_{j=1}^{p} a_{i,j}}\int_0^1 \operatorname{sgn}\left(\widetilde{W}^{\mu}_{\eta_i}(s)\right)W(\eta_i(s))\,\mathrm{d}s. \tag{6}$$

To conclude about the convergence of the expectation of the l.h.s. to the expectation on the r.h.s. of (6), note that the sequence $\frac{1}{T^{1.5}}\sum \operatorname{sgn}\left(\widetilde{x}^{\mu}_{i,t-1}\right)x_{i,t-1}$ is uniformly $L_2$-bounded (in $t$ as well as in $i$) and as such uniformly integrable so convergence of the expectations holds, as required for the result.

## Proof of Lemma 2

Since $\left\|\boldsymbol{S}_T - \overline{\boldsymbol{\Omega}}\right\| \le \left\|\boldsymbol{S}_T - \widehat{\boldsymbol{\Sigma}}_{\varepsilon}\right\| + \left\|\widehat{\boldsymbol{\Sigma}}_{\varepsilon} - \overline{\boldsymbol{\Omega}}\right\|$, we only have to prove that $\left\|\boldsymbol{S}_T - \widehat{\boldsymbol{\Sigma}}_{\varepsilon}\right\| = O_p\left(NT^{-0.5}\right)$ thanks to (4). With $\left\|\boldsymbol{I}\right\| = 1$, $\left\|\boldsymbol{S}_T - \widehat{\boldsymbol{\Sigma}}_{\varepsilon}\right\| \le |\kappa_{1T}| + |\kappa_{2T} - 1|\left\|\widehat{\boldsymbol{\Sigma}}_{\varepsilon}\right\|$. Since $\left\|\widehat{\boldsymbol{\Sigma}}_{\varepsilon}\right\| = O_p(N)$, $|\kappa_{2T} - 1| = b_T^2/d_T^2$ and $m_T = \frac{1}{N}\operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}_{\varepsilon}\right) = \frac{1}{NT}\sum\sum_{i=1}^{N}\overline{\varepsilon}_{it}^2 = O_p(1)$, proving $\left\|\boldsymbol{S}_T - \widehat{\boldsymbol{\Sigma}}_{\varepsilon}\right\| = O_p\left(NT^{-0.5}\right)$ reduces to showing

$$\frac{b_T^2}{d_T^2} = O_p\left(T^{-0.5}\right).$$

An upper bound for $b_T^2$ is derived as follows. Recall, $b_T^2 = \min(\bar{b}_T^2, d_T^2)$ with $\bar{b}_T^2 = \left(\sum\left(\overline{\boldsymbol{\varepsilon}}_t'\overline{\boldsymbol{\varepsilon}}_t/T\right)^2 - \operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^2\right)/T\right)/N$. Due to the symmetry of $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$, the trace on the r.h.s. amounts to the sum of the squared elements of $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$; the elements have uniformly bounded variance (cf. the proof of Lemma 1), so the squares have uniformly bounded expectation and thus $\frac{1}{T}\operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^2\right) = O_p\left(N^2T^{-1}\right)$. It then follows analogously that $\left(\overline{\boldsymbol{\varepsilon}}_t'\overline{\boldsymbol{\varepsilon}}_t/T\right)^2 = O_p\left(N^2/T^2\right)$ and thus $\bar{b}_T^2 = O_p\left(NT^{-1}\right)$. Since $b_T^2 = \min(\bar{b}_T^2, d_T^2)$, we need a lower bound for $d_T^2$. The trace involved in the expression of $d_T^2$ amounts to the sum of squared elements of $\widehat{\boldsymbol{\Sigma}}_{\varepsilon} - m_T\boldsymbol{I}$. Due to the $\sqrt{T}$-consistency of the elements of $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$, we have

$$d_T^2 = \operatorname{tr}\left(\left(\overline{\boldsymbol{\Omega}} - m_T\boldsymbol{I}\right)\left(\overline{\boldsymbol{\Omega}} - m_T\boldsymbol{I}\right)'\right)/N + O_p\left(NT^{-0.5}\right).$$

The $O_p$ term on the right-hand side indicates an upper bound, though, so we derive the desired lower bound for $d_T^2$ from the behavior of $\overline{\boldsymbol{\Omega}} - m_T\boldsymbol{I}$. In fact is is sufficient to examine $\overline{\boldsymbol{\Omega}}$, given that $m_T = O_p(1)$. Under Assumption 3 (a), $\operatorname{tr}\left(\overline{\boldsymbol{\Omega}}^2\right)$ is of exact magnitude order $N^2$, so $\operatorname{tr}\left(\left(\overline{\boldsymbol{\Omega}} - m_T\boldsymbol{I}\right)\left(\overline{\boldsymbol{\Omega}} - m_T\boldsymbol{I}\right)'\right)/N$ is bounded away from 0 (it is in fact of order at least $N$) and $b_T^2 = O_p(N/T)$. Using again the fact that $d_T^2$ is of order at least $N$, we obtain that

$$\frac{b_T^2}{d_T^2} = O_p\left(\frac{1}{T}\right),$$

which is sufficient for the result.

## References


Bai, J. and S. Ng (2004). A panic attack on unit roots and cointegration. *Econometrica 72*(4), 1127–1177.

Boswijk, H. P. (2005). Adaptive testing for a unit root with nonstationary volatility. UvA-Econometrics Discussion Paper 07, Universiteit van Amsterdam.

Breitung, J. and S. Das (2005). Panel unit root tests under cross sectional dependence. *Statistica Neerlandica 59*(4), 414–433.

Cavaliere, G. (2004). Unit Root Tests under Time-Varying Variances. *Econometric Reviews 23*(3), 259–292.

Cavaliere, G. and A. M. R. Taylor (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics 140*, 919–947.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Demetrescu, M. and C. Hanck (2011). IV estimators for autoregressive processes under nonstationary error volatility. *mimeo*.

Demetrescu, M., U. Hassler, and A.-I. Tarcolea (2006). Combining significance of correlated statistics with application to panel data. *Oxford Bulletin of Economics and Statistics 68*(5), 647–663.

Hanck, C. (201x). An intersection test for panel unit roots. *Econometric Reviews*.

Harris, D., D. I. Harvey, S. J. Leybourne, and N. D. Sakkas (2010). Local asymptotic power of the Im-Pesaran-Shin panel unit root test and the impact of initial observations. *Econometric Theory 26*, 311–324.

Hartung, J. (1999). A note on combining dependent tests of significance. *Biometrical Journal 41*, 849–855.

Im, K. S., M. H. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics 115*(1), 53–74.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*, 365–411.

Lütkepohl, H. (1996). *Handbook of Matrices*. New York: John Wiley & Sons.

Moon, H. R. and B. Perron (2004). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics 122*, 81–126.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics 22*(2), 265–312.

Shin, D. W. and S. Kang (2006). An Instrumental Variable Approach for Panel Unit Root Tests under Cross-Sectional Dependence. *Journal of Econometrics 134*(1), 215–234.

Shin, D. W. and O. Lee (2001). Tests for Asymmetry in Possibly Nonstationary Time Series Data. *Journal of Business and Economic Statistics 19*(2), 233–244.

Shin, D. W. and O. Lee (2003). An instrumental variable approach for tests of unit roots and seasonal unit roots in asymmetric time series models. *Journal of Econometrics 115*(1), 29–52.

So, B. S. and D. W. Shin (1999). Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals. *Econometric Theory 15*(2), 165–176.

Stock, J. H. and M. W. Watson (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*(1), 159–218.